Structure, Vol. 9, 513–525, June, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00612-8

The κ-carrageenase of *P. carrageenovora* Features a Tunnel-Shaped Active Site: A Novel Insight in the Evolution of Clan-B Glycoside Hydrolases

Gurvan Michel,[1,2] Laurent Chantalat,[1] Emile Duee,[1] Tristan Barbeyron,[2] Bernard Henrissat,[3] Bernard Kloareg,[2] and Otto Dideberg[1,4]

[1] Laboratoire de Cristallographie Macromoléculaire
Institut de Biologie Structurale Jean-Pierre Ebel
CNRS/CEA
41 Avenue des Martyrs
38027 Grenoble Cedex 1
France

[2] Station Biologique de Roscoff
UMR 1931 (CNRS and Laboratoires Goëmar)
Place Georges Teissier, BP 74
29682 Roscoff Cedex
France

[3] Architecture et Fonction des Macromolécules Biologiques
CNRS-IFR 1, 31
Chemin Joseph Aiguier
13402 Marseille Cedex 20
France

Summary

Background: κ-carrageenans are gel-forming, sulfated 1,3-α-1,4-β-galactans from the cell walls of marine red algae. The κ-carrageenase from the marine, gram-negative bacterium *Pseudoalteromonas carrageenovora* degrades κ-carrageenan both in solution and in solid state by an endoprocessive mechanism. This β-galactanase belongs to the clan-B of glycoside hydrolases.

Results: The structure of *P. carrageenovora* κ-carrageenase has been solved to 1.54 Å resolution by the multiwavelength anomalous diffraction (MAD) method, using a seleno-methionine-substituted form of the enzyme. The enzyme folds into a curved β sandwich, with a tunnel-like active site cavity. Another remarkable characteristic is the presence of an arginine residue at subsite −1.

Conclusions: The crystal structure of *P. carrageenovora* κ-carrageenase is the first three-dimensional structure of a carrageenase. Its tunnel-shaped active site, the first to be reported for enzymes other than cellulases, suggests that such tunnels are associated with the degradation of solid polysaccharides. Clan-B glycoside hydrolases fall into two subgroups, one with catalytic machinery held by an ancestral β bulge, and the other in which it is held by a regular β strand. At subsite −1, all of these hydrolases exhibit an aromatic amino acid that interacts with the hexopyranose ring of the monosaccharide undergoing catalysis. In addition, in κ-carrageenases, an arginine residue recognizes the sulfate-ester substituents of the β-linked κ-carrageenan monomers. It also appears that, in addition to the nucleophile and acid/base catalysts, two other amino acids are involved with the catalytic cycle, accelerating the deglycosylation step.

[4] Correspondence: otto@ibs.fr

Introduction

Agars and carrageenans are hydrophilic polysaccharides with unique rheological properties and are widely used as texturing and moisturizing agents in various industries [1]. They are extracted from the cell walls of marine red algae (Rhodophyceaea), a eukaryotic lineage as ancient as one billion years [2], where they naturally occur in the form of pseudocrystalline matrix polysaccharides associated with cellulose [3]. They consist of a linear backbone of galactopyranose units linked by alternating α(1 → 3) and β(1 → 4) linkages. While all of the β-linked residues of these galactans are in the D configuration, the α(1,4)-linked galactose units are in the L configuration in agars and in the D configuration in carrageenans. A further layer of complexity is introduced by the occurrence of a 3,6-anhydro bridge in the α(1,4)-linked galactose residues and, particularly in carrageenans, by a variable number of sulfate substituents per disaccharide repeating unit [3].

As shown from conformational analyses and X-ray diffraction studies [4, 5, 6], the succession of α(1 → 3) and β(1 → 4) linkages as well as the occurrence of 3,6-anhydro-D-galactose bridges in the polysaccharide backbone favor the formation of twisted ribbons, with alternating $^1C_4$ and $^4C_1$ chair conformations for the galactopyranose rings (Figure 1) and an overall 3_1 symmetry. Thus, carrageenans readily adopt double-helical conformations, leading to the formation of aggregates of double-stranded helices, which are further stabilized by counter-ions such as potassium and calcium [7]. However, helix propagation is interrupted by the occurrence of natural structural discontinuities or "kinks", resulting from the occasional replacement of 3,6-anydro-D-galactose residues by D-galactose-6-sulfate or D-galactose-2,6-disulfate moieties. Self-association of these polysaccharides into large crystalline clusters is thus prevented, and, instead, they form three-dimensional solid networks in aqueous solutions consisting of crystalline junction zones of double helix aggregates separating amorphous zones of single-stranded helices and coils [4].

We maintain various marine bacteria that secrete κ-, ι-, and λ-carrageenases. These bacteria all belong to the genera *Pseudoalteromonas* or *Alteromonas* and to the genus *Zobellia*, two groups of phylogenetically distant bacteria. These hydrolases cleave the internal β(1 → 4) linkages of carrageenans, yielding oligogalactans of the neocarrabiose series [8, 9]. In the case of *Pseudoalteromonas carrageenovora* κ-carrageenase, hydrolysis of the substrate is accompanied by a rapid fall in specific

Key words: glycoside hydrolase; MAD; sulfated galactan; X-ray structure

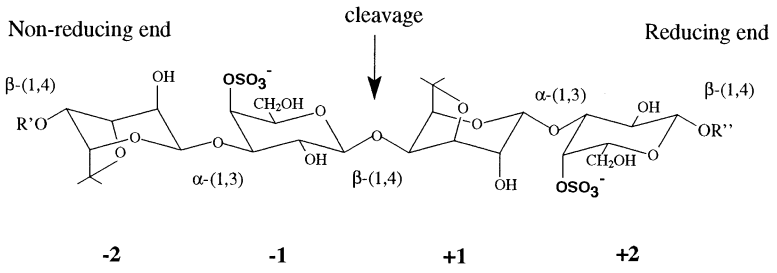

Figure 1. Schematic Diagram Showing Two Disaccharide-Repeating Units of κ-carrageenan

The two different glycosidic bonds in κ-carrageenan are labeled, and further extensions of the polymer are indicated by R′ and R″ at the nonreducing and reducing ends, respectively. As κ-carrageenase cleaves the β-(1,4) glycosidic bond, its subsites for substrate binding are labeled in accordance to the established nomenclature [55].

viscosity and a marked increase in reducing power, indicating an endoattack mechanism. The two major end products are neocarrabiose-sulfate and neocarratetraose-sulfate [8], and they appear as soon as within one minute of incubation [10], suggesting a processive character for this enzyme. Carrageenase products were recently recognized as virulence factors for pathogens of red algae, with the sulfate susbstituents acting as a main factor in host recognition [11].

Carrageenases provide the opportunity to investigate the structure-function relationships of the hydrolases that degrade self-associating sulfated polysaccharides, and we have undertaken their molecular analysis. The gene products of the κ-carrageenase of *P. carrageenovora* and those of another marine bacterium, *Zobellia galactanovorans*, have been described in detail [12, 13]. Even though κ-carrageenan differs from agarose by the D configuration of the α-linked galactose residues and by the presence of one sulfate substituent at C4 on the β-linked D-galactose residues (Figure 1), κ-carrageenases were shown to belong to the glycoside hydrolases family-16, along with several β-agarases. As other members of this family [14], the κ-carrageenase from *P. carrageenovora* was shown to proceed with an overall retention of the anomeric configuration and to exhibit transglycosylating activity [15].

Family-16 of the glycoside hydrolases [16] also comprises 1,3-β-glucanases (laminarinases), 1,3-1,4-β-glucan-hydrolases (also known as lichenases), and xyloglucan endotransglycosylases (XETs) [13]. Based on phylogenetical evidence, we proposed that all of these enzymes share a common ancestor that has laminarinase activity. Even though these proteins have now diverged significantly in their primary sequences, they all feature a common catalytic motif, (E[ILV]D[IVAF][VILMF](0,1)E). A detailed crystallographic analysis of the active sites of *Bacillus licheniformis* and *B. macerans* lichenases [17, 18] as well as site-directed mutagenesis experiments [19, 20] have shown that the two glutamic acid residues in the conserved motif are the nucleophile and the general base involved in catalysis, whereas the aspartic acid residue is important in maintaining the relative position of these catalytic amino acids. These three amino acid residues are also strictly conserved in glycoside hydrolases from family-7, suggesting that the two families form a "clan", referred to as clan-B of glycoside hydrolases [21, 22].

As members of family-16 glycoside hydrolases, κ-carrageenases should display a general fold similar to that previously determined for *B. licheniformis* and *B. macerans* lichenases [17, 18]. However, κ-carrageenases and lichenases share only 20% overall identity in their primary structures. Even in their catalytic motif, κ-carrageenases depart from lichenases by the presence of one additional amino acid, an insertion that could affect the active-site fine topology. A high-resolution three-dimensional structure of a κ-carrageenase was therefore required to define the amino acid residues involved in the recognition and cleavage of κ-carrageenans. As a preliminary step toward this goal, we recently reported the expression, purification, and crystallization of the κ-carrageenase from *P. carrageenovora* [23].

Here, we report the crystal structure of *P. carrageenovora* κ-carrageenase, to our knowledge the first three-dimensional structure of a carrageenase. Reminiscent of the processive cellulases known as cellobiohydrolases, *P. carrageenovora* κ-carrageenase features a tunnel-shaped active site, suggesting that the acquisition of such an active-site topology is a convergent evolution process, characteristic of polysaccharidases capable of degrading solid substrates. This structure also provides a novel insight into the evolution of clan-B glycoside hydrolases. We show, in particular, that they fall into two subgroups. The first group includes the family-16 κ-carrageenases, agarases, laminarinases, as well as the family-7 cellulases and has an ancestral, β-bulged catalytic site. The other group comprises the more recent lichenases and xyloglucan endotransglycosylases and displays a regular, β-stranded catalytic center with one amino-acid deletion. We also propose that the conspicuous aspartic acid residue of the catalytic site of clan-B glycoside hydrolases as well as a histidine residue conserved in κ-carrageenases, β-agarases, laminarinases, and family-7 cellulases are involved through proton trafficking in the deglycosylation step of the catalytic cycle.

Results and Discussion

Overall Structure of *P. carrageenovora* κ-carrageenase

The three-dimensional crystal structure of *P. carrageenovora* κ-carrageenase was determined at 1.74 Å resolution by the multiwavelength anomalous diffraction (MAD) method, using a crystal of a seleno-methionine-substituted form of the enzyme. The resulting experimental electron density map was of high quality (Figure 2). The model was then refined at 1.54 Å resolution to a crystallographic R factor of 17.9% and an R_{free} of 19.3%. Crystallographic statistics are reported in Table 1. The asymmetric unit encompasses amino acids 27–297 of κ-carrageenase, 7 cadmium ions, 7 chloride ions, and

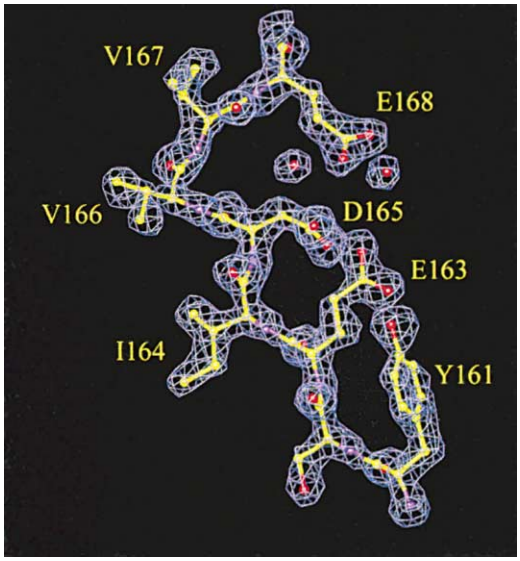

Figure 2. Solvent-Flattened MAD Electron Density Map at 1.74 Å Resolution Contoured at 1.5 σ of the Active Site

Water molecules are marked with red spheres. The protein's atoms are shown in red, blue, and yellow for oxygen, nitrogen, and carbon atoms, respectively. E163 and E168 are the nucleophile and the acid/base catalyst, respectively. This figure was created using O [50].

405 water molecules. Residues 26 and 298–301 are not visible in the ($2F_o$-F_c) electron density map and thus are presumed to be in a disordered conformation.

The *P. carrageenovora* κ-carrageenase is an almost all-β protein with a globular shape and dimensions of 60 × 50 × 40 Å. It consists of a β sandwich formed by two main, closely packed and curved antiparallel β sheets of six and seven strands each, creating a deep channel (Figure 3). The interface between these β sheets, referred to as A and B, is mainly hydrophobic, but it also presents several hydrogen bond networks that link the internal and external sheets (H180 ↔ N206 and Q241, Q257 ↔ T85 and T284, W57 ↔ D47, D165 ↔ W232, H205 and Q202 ↔ K242, and H183 ↔ W247). According to the secondary structure assignment by DSSP [24], three minor antiparallel β sheets, A1, C, and D, and a one-turn α helix complete this fold. The long and highly twisted strands β5 and β6 (β sheet A1) are closely associated to the strands β19 and β20 (β sheet D), topping the channel. Their mostly hydrophobic interactions are strengthened at the tip of the strands by several structural features. The only disulfide bridge in the protein, between C98 and C268, connects the turns between β5 and β6 and β19 and β20, respectively, while the Q270 side chain binds strand β5 through a bidentated hydrogen bond with the backbone carbonyl group and nitrogen atom of W95 and D96, respectively. A carboxylate group of this latter residue also presents a short hydrogen bond (2.51 Å) with the hydroxyl group of Y272. These extensions of the external sheet and the strands β12 and β13 of the internal sheet concur in creating a long tunnel in the concave face of the protein (see also Figure 6).

Five cadmium ions, identified in an anomalous Fourier difference map using data collected in our laboratory at 1.5418 Å (data not shown), are bound to the surface acidic amino acid residues, tightening the crystal packing. Another cadmium ion is bound to H183 in the active site, but this is likely to be a crystallization artifact. Finally, a seventh cadmium ion is bound with an almost octahedral geometry on the convex face of the enzyme. It is coordinated by the carbonyl oxygen atoms of E48, G79, and D289, a carboxylate oxygen atom of this latter residue, and three water molecules. The crystal structure also displays seven chloride ions that are ionically bound to the cadmium ions.

As expected from its structural membership to family-16 hydrolases [12, 13], the fold of *P. carrageenovora* κ-carrageenase is more closely related to other family-16 enzymes than to family-7 members. To compare those three-dimensional structures, the DALI server was used [25]. Out of the 271 residues of κ-carrageenase, a core of 198 residues can be superimposed with the structure of *B. macerans* 1,3-1,4-β-glucanase (PDB code 2ayh), with a root-mean-square distance (rmsd) on Cα atoms of 2.4 Å (Figure 4). Similarly, the structure of the κ-carrageenase partially overlays that of *Fusarium oxysporum* endoglucanase I [26], such that 185 residues have an rmsd of 3.5 Å.

Active Site Structure

Catalytic Nucleophile and Acid/Base Residues

Glycosidic bond cleavage by retaining glycosidases takes place via a double-displacement mechanism involving two catalytic carboxyl groups [22, 27, 28]. In the first step (glycosylation), one of the carboxyl group acts first as a general acid, protonating the glycosidic oxygen concomitantly with bond cleavage. The other catalytic residue acts as a nucleophile, forming a covalent glycosyl-enzyme intermediate. In the second step (deglycosylation), the general acid now acts as a general base and deprotonates a water molecule, which attacks the anomeric carbon and releases the sugar. Both steps occur via transition states with substantial oxocarbenium ion character [27].

The nucleophile and acid/base catalytic residues of family-16 of glycoside hydrolases have already been unambiguously identified in the *B. macerans* 1,3-1,4-β-glucanase [17, 19, 20]. The corresponding residues in *P. carrageenovora* κ-carrageenase are E163 (nucleophile) and E168 (acid/base) on strand β10. However, compared to the homologous β strand that, in lichenases, carries the catalytic residues, one valine residue is inserted in the κ-carrageenases [12, 13]. Yet, in spite of this insertion, the catalytic residues of the two enzymes are exactly superimposed, because of the occurrence, in strand β10 of κ-carrageenase, of a β bulge that is absent in the lichenase (Figure 2). The κ-carrageenase hydrophobic core has undergone some changes to accommodate this β bulge. In the case of *B. macerans* 1,3-1,4-β-glucanase, residue I108, which points toward the protein hydrophobic core, interacts with F110 and F149. In *P. carrageenovora* κ-carrageenase, the replacement of the former residue by the V166–V167 β bulge is compensated by the substitution of the two phenylalanine residues by less voluminous amino acids, L169 and

Table 1. Data Reduction, Phasing, and Refinement Statistics

	Peak	Edge	Remote
Data Collection			
Wavelength (Å)	0.9755	0.9787	0.8860
f′, f″	−7.5, 5.6	−9.2, 4.0	−2.1, 3.6
Resolution (Å)[a]	1.74 (1.80–1.74)	1.74 (1.80–1.74)	1.54 (1.60–1.54)
Total data	84,437	86,479	120,447
Unique data	26,814	27,031	38,205
Redundancy	3.15	3.20	3.15
Completeness (%)	99.3 (97.2)	99.2 (95.5)	97.7 (81.3)
$I/\sigma(I)$	15.3 (5.8)	19.4 (8.0)	16.1 (6.2)
R_{sym} (%)[b]	2.9 (7.0)	2.9 (7.3)	3.0 (11.6)
Phasing (SOLVE)			
R_{ano} (%)[c]	6.1	5.0	4.7
R_{iso} (%)[d]	4.5	5.5	
FOM[e] before/after DM (to 1.54 Å)	0.68/0.84		
Refinement			
Resolution range (Å)			24.25–1.54
Number of unique reflections			38,205
R_{work} (R_{free})[f]			17.9 (19.3)
Rmsd[g] bonds (Å)			0.005
Rmsd angles (°)			1.4
Rmsd in main chain bond B factor (Å^2)			0.28 (1.5)[h]
Rmsd in side chain bond B factor (Å^2)			0.37 (2.0)
Quality of Ramachandran plot[i]			
Percentage of residues in most favored regions			86.1
Percentage of residues in additional allowed regions			13.5
Percentage of residues in generously allowed regions			0.4
Number of atoms (average B values [Å^2])			
Protein			2,229 (10.3)
Water			405 (24.1)
Ion (Cd-Cl)			14 (16.3)

[a] Values in parentheses correspond to the highest resolution shell.
[b] $R_{sym} = \Sigma|I - I_{av}|/\Sigma I$, where the summation is over all symmetry equivalent reflections.
[c] $R_{ano} = \Sigma|I(+) - I(-)|/\Sigma|I_{av}|$ for anomalous differences, where I_{av} is the average of Friedel amplitudes at a single wavelength.
[d] $R_{iso} = \Sigma|I_{\lambda i} - I_{\lambda j}|/\Sigma|I_{av\lambda}|$ for dispersive differences, where $I_{av\lambda}$ is the average amplitude at two wavelengths.
[e] FOM, figure of merit.
[f] R calculated on 5% of data excluded from refinement.
[g] Rmsd, root-mean-square deviation.
[h] Target values.
[i] Figures from PROCHECK [59].

V223. Considering only these substitutions, an excess of 27 Å^3 can be calculated, which results from the valine insertion in κ-carrageenase. Slight movement in the protein structure can accommodate this excess. Interestingly, while β bulges are thought to play an important biological role in proteins [29], such a role is not observed in family-16 glycoside hydrolases.

Asp165 and His183 Are Involved with the Catalytic Cycle

As expected for a retaining glycoside hydrolase [22], the carboxyl groups of the nucleophile (E163) and the acid/base catalyst (E168) amino acids of *P. carrageenovora* κ-carrageenase are separated by 5.4 Å. E168 is involved in hydrogen bonding with four solvent molecules only, which are the candidates to act as catalytic incoming water molecules. In contrast, E163 is firmly oriented by hydrogen bonds to Y161, which is conserved in the two κ-carrageenases known so far, and to D165. This latter residue, which is indeed strictly conserved throughout the clan-B of glycoside hydrolases (see Figure 8), makes a short hydrogen bond (2.54 Å) with the nucleophile E163 OE2. Such a small distance suggests that this is a low-barrier hydrogen bond [30] and that the proton can move freely between E163 OE2 and D165 OD1. Mutation of the aspartate residue to asparagine in *B. macerans* 1,3-1,4-β-glucanase results in only residual activity, less than 1% of wild-type [20]. It was suggested that this conserved aspartic acid plays an important, but not crucial, role in catalysis. However, the D → N mutation does not abolish the possibility of a hydrogen bond between E103 and the N105 amide group. In *Trichoderma reesei* endoglucanase I, the equivalent residue, D199, is believed to maintain the nucleophile, E197, in the negatively charged state at the beginning of the catalytic cycle and probably the correct pK_a for E197 and E202 [31]. As illustrated in Figure 5, we propose an additional role for this strictly conserved amino acid. In the catalytic cycle, after the formation of the glycosyl-

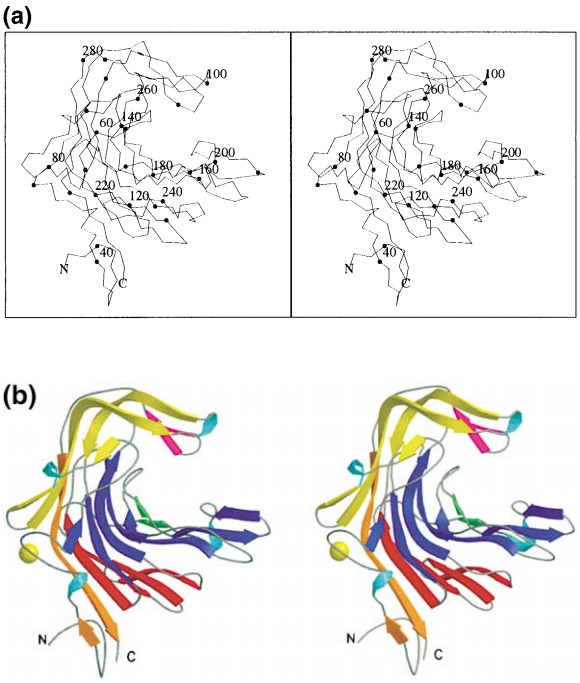

Figure 3. Folding of the *P. carrageenovora* κ-carrageenase

(a) Stereo view of the Cα trace of the protein. The N terminus, C terminus, and every 20 residues are labeled; every 10 residues are marked with a black dot.

(b) Stereo view of the ribbon representation of the structure. The five β sheets A (β8-15-16-17), A1 (β3-4-5-6), B (β2-7-18-9-10-11-14), C (β12-13), and D (β19-20) are shown in red, yellow, blue, green, and magenta, respectively. The β1 and β21, which belong to both β sheets A and A1, are shown in orange; short helices are shown in light blue. The yellow sphere represents a cadmium ion that occupies the conserved calcium binding site.

Figures 3, 4, and 7 were prepared with Molscript [56].

enzyme intermediate, the low-barrier hydrogen bond is no longer possible between E163 and D165 (numbering of κ-carrageenase), leading to the protonation of the aspartic acid residue. In the transition state, the D165 proton competes with the positive charge of the anomeric C1 atom, lowering the strength of the covalent linkage between the anomeric carbon and E163, hence favoring the reformation of a low-barrier hydrogen bond

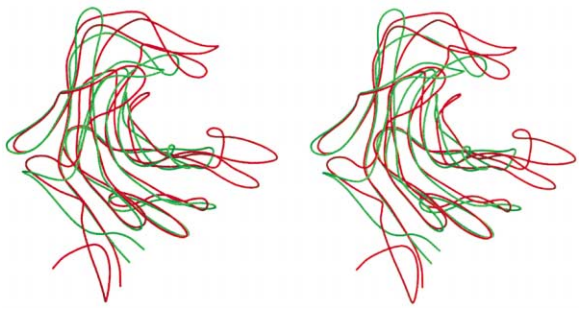

Figure 4. Stereo View of *P. carrageenovora* κ-carrageenase Superimposed to *B. macerans* 1,3-1,4-β-glucanase

κ-carrageenase and 1,3-1,4-β-glucanase are shown as red and green coils, respectively.

with OE2 of this latter residue. D165 may thus be involved in catalysis by accelerating the deglycosylation step, possibly by two orders of magnitude, as suggested by the low residual activity of the mutated enzyme [20]. This hypothesis is supported by a close inspection of the structure of *B. macerans* 1,3-1,4-β-glucanase in complex with a suicide inhibitor (1byh) compared to the native enzyme structure (2ayh) [17]. In the inhibitor complex, the distance between D105 OD1 and E103 OE2 is increased from 2.50 to 3.54 Å, which precludes any hydrogen bonding between these two residues in the glycosyl-enzyme intermediate state. Furthermore, in the case of κ-carrageenase, D165 OD1 is also hydrogen-bonded to H183. This histidine residue appears to be conserved in β-agarases and 1,3-β-glucanases as well as in family-7 cellulases (Figure 8). It is thought to be involved in proton trafficking, thus ensuring regeneration of the correct protonated states of the catalytic residues [31]. We suggest that both the conserved aspartic acid and histidine residues (D165 and H183, numbering of κ-carrageenase) cooperate in proton trafficking during the deglycosylation step.

P. carrageenovora κ-*carrageenase Features a Closed, Tunnel-like Active Site*

As emphasized by the representation of the electrostatic surfaces of *P. carrageenovora* κ-carrageenase and *B. macerans* 1,3-1,4-β-glucanase (Figure 6), another striking difference between these two proteins is the presence of a tunnel-like active site in the κ-carrageenase. The tunnel is formed by the extension of β strands that are also present in the lichenase, such as strands β5 and β6 (from S89 to Y109), or of loops, such as those between strands β9 and β10 (from D150 to Q160), as well as by the occurrence of β strands that are unique to the κ-carrageenase, such as strands β12 and β13 (from I185 to W195) and β19 and β20 (from L259 to G281).

Finally, relative to 1,3-1,4-β-glucanases, two other structural modifications alter the fine topology of the κ-carrageenase active site. First, between the κ-carrageenase strands β11 and β10 (the latter carrying the catalytic residues), a glycine residue is substituted by Q171, a change which occurs in κ-carrageenases only (see Figure 8). The mutation is associated with a 6 Å movement of the T170–V175 loop toward the active site compared with the position of the equivalent loop in 1,3-1,4-β-glucanases. The side chain of Q171 is thus positioned near the catalytic residue E168, and it takes part in a hydrogen bond network between various residues conserved in κ-carrageenases as follows. Q171 is hydrogen-bound through its NE2 to D179 OD2, which also makes a salt bridge with K172, and through its OE1 to the NH2 and NE of R260, an arginine residue likely to be critical for κ-carrageenan recognition (see below). Second, the κ-carrageenase markedly differs from *B. macerans* 1,3-1,4-β-glucanase in the structure of the loop between the β2 and β3 strands (numbering of κ-carrageenase). In the lichenase, this loop (A21–N26) carries a tyrosine residue, Y24, involved with enzyme specificity: it is thought to interact with a mixed-linked β-hexaglucoside through stacking against a glucose ring in subsite −4 and through hydrogen bonding to the hydroxyl O4 of the next glucose moiety in subsite −3 [32]. In *P. carra-*

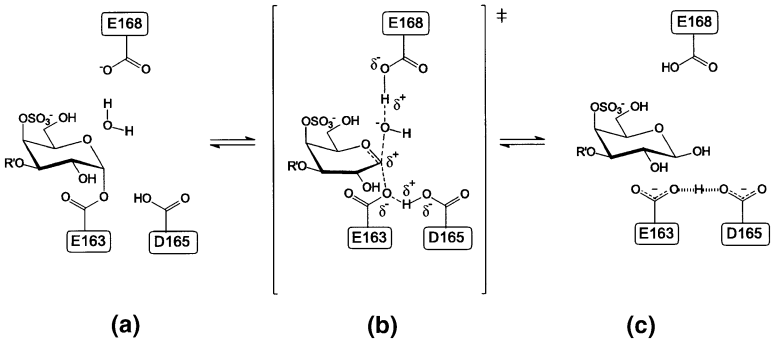

Figure 5. Possible Role of D165 in the Reaction Catalyzed by κ-carrageenase

Only the saccharide in subsite −1 and three acidic amino acid residues involved in the catalytic process are represented. The first step (glycosylation) is not represented.

(a) The glycosyl-enzyme intermediate. Since the nucleophile E163 is covalently bound to the anomeric carbon C1 of the sugar moiety, the low-barrier hydrogen bond between E163 and D165 is not possible.

(b) The oxocarbonium transition state. The D165 proton competes with the positively charged anomeric carbon to form a hydrogen bond with E163, accelerating the deglycosylation step.

(c) The reaction product. The low-barier hydrogen bond between E163 and D165 is restored, and the nucleophile is maintained in a negatively charged state for the next catalytic cycle.

geenovora κ-carrageenase, the equivalent loop enters (by 5–6 Å) the active site cleft more deeply, allowing the carboxylate oxygen OE2 of E62 to establish a short (2.5 Å), strong hydrogen bond with the hydroxyl group of Y145. These two latter residues thus completely obstruct the subsite −3 identified in the 1,3-1,4-β-glucanases. Altogether, compared to lichenases, these two loop modifications result in a shrinking and bending of the catalytic cleft of κ-carrageenases.

Molecular Bases of κ-carrageenan Recognition

Model of the Substrate into the Protein Channel

The small size of the catalytic tunnel of κ-carrageenase cannot accommodate a double helix of κ-carrageenan. It follows that the enzyme can attack κ-carrageenan chains either as single-stranded helices or as random coils. In an attempt to identify the crucial residues for κ-carrageenan recognition, single-stranded κ-carrageenan oligomer models were therefore docked into the active-site tunnel of the κ-carrageenase. The structure of the family-7 *F. oxysporum* endoglucanase I complexed with a nonhydrolyzable thiopentasaccharide, referred to as thio-DP5 [26], was used as a guide to position the substrate in the κ-carrageenase catalytic site. As endoglucanase I and κ-carrageenan both hydrolyze β-1,4-linkages with a conserved catalytic machinery, the D-galactose-4-sulfate unit of κ-carrageenan must overlay the D-glucose residue of thio-DP5 at the site of cleavage, thus determining the positioning of κ-carrageenan in subsite −1 of κ-carrageenase.

However, when the substrate is docked in its ordered, extended conformation [6], not only the 3,6-anhydro-D-galactose residue, at subsite +1, but also the subsequent D-galactose residues located toward the reducing end sterically clash with the protein, for any value given to the torsion angles ϕ (H1-C1-O4-C4) and Ψ (C1-O4-C4-H4) of the β-(1,4) bond. In contrast, if, as observed for thio-DP5, the D-galactose-4-sulfate residue in subsite −1 adopts a distorted $^{1,4}B$ boat conformation, the glycosidic moieties at the reducing end no longer make forbidden contacts. The fine structure of subsites −1 and +1 of κ-carrageenase is most likely responsible for the conformational change of D-galactose-4-sulfate into a distorted $^{1,4}B$ boat, which is close to the conformation of the transition state [26].

At the nonreducing end, because of its 3_1 symmetry, the κ-carrageenan strand does not follow the protein channel, but appears instead to coil around the side chains of the residues forming the tunnel. Thus, unlike 1,3-1,4-β-glucanases and cellulases [32, 33], the substrate chain does not come out of the catalytic cavity parallel to the internal β sheet, but almost perpendicular to it. Such a topology is consistent with the bending of the catalytic cleft discussed above. Based on these modeling studies, seven subsites, five before the site of cleavage and two after, for the accomodation of the D-galactosyl units of κ-carrageenan can be identified in the κ-carrageenase tunnel (Figure 7). Subsites −5, −3, and +2 and subsites −4, −2, and +1 accommodate three D-galactose-4-sulfate residues and three 3,6-anhydro-D-galactose units, respectively, which maintain their ground state original $^4C_1$ and $^1C_4$ chair conformations (Figure 1), while subsite −1 binds one D-galactose-4-sulfate unit in a modified, distorted $^{1,4}B$ boat conformation.

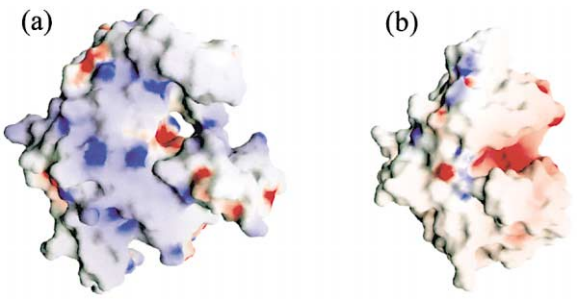

Figure 6. Molecular Surface Colored According to Electrostatic Potential Ranging from Deep Blue, +, to Red ,−

(a) κ-carrageenase.

(b) 1,3-1,4-β-glucanase; the orientation is the same as in Figure 4. This figure was created with Grasp [57].

Ionic and Nonpolyelectrolyte Protein-Substrate Interactions

Since κ-carrageenan is a polyanion harboring one sulfate-ester group per disaccharide-repeating unit, it is

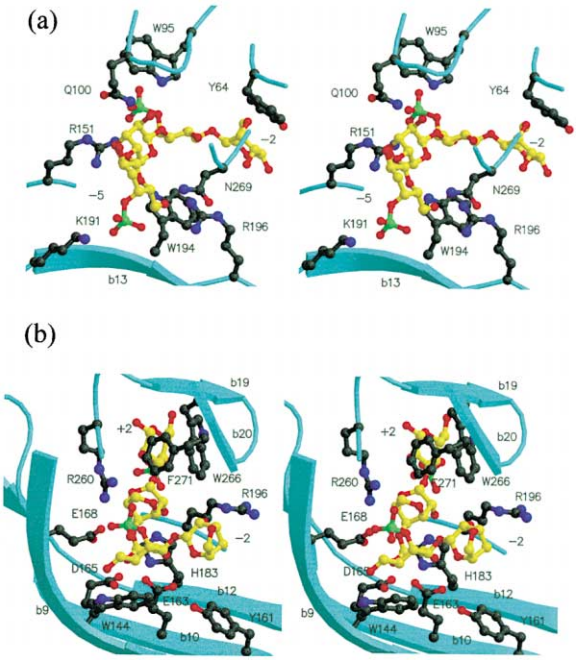

Figure 7. Stereo View of the κ-carrageenan Model in the Active Site of the κ-carrageenase

(a) Subsites −5 to –2.

(b) Subsites −2 to +2.

κ-carrageenase strands are labeled, important residues are shown as balls and sticks, nitrogens are shown in blue, oxygens are shown in red, and carbons are shown in black. The substrate is represented in balls and sticks, sulfurs are shown in green, oxygens are shown in red, and carbons are shown in yellow.

likely that substrate binding involves ionic interactions with basic amino acid residues. Consistently, the electrostatic surface of κ-carrageenase exhibits a markedly higher density of positive patches than that of *B. macerans* 1,3-1,4-β-glucanase (Figure 6). On the other hand, the 3,6-anhydro-D-galactose residues of κ-carrageenan are neutral and bear only one free hydroxyl group. Hence, they are more hydrophobic than D-galactose-4-sulfate or even regular D-galactose residues, and the part of hydrophobic interactions of κ-carrageenase with its substrate should be significant for the binding of the neutral sugar moieties. Altogether, one might expect that substrate binding by κ-carrageenase involves ionic interactions between the polysaccharide sulfate-ester substituents and the protein basic residues and a specific hydrophobic environment for the 3,6-anhydro-D-galactose units.

The ionic and hydrophobic interactions involved with κ-carrageenan recognition were analyzed in more detail on the basis on the enzyme-substrate complex model proposed above (Figure 7). In subsites −5 and −4, the D-galactose units are outside the tunnel, near the β13 strand, which extends out toward the solvent. In subsite −5, the O6 hydroxyl group of D-galactose-4-sulfate is hydrogen-bound to N269 NE2 and R196 NH1 and NH2, two conserved residues in κ-carrageenases. The glutamine residue Q100 NE2 could also form a hydrogen bond with the O2 hydroxyl group of this galactopyranose ring, assuming a rotation of 90° of the amino acid side chain. A salt bridge between the sulfate ester group and K191 completes the binding of this D-galactose-4-sulfate residue. In spite of its relatively more hydrophobic character, the 3,6-anhydro-D-galactose residue in subsite −4 is involved in hydrogen bonding with two tryptophans, the O2 hydroxyl group giving its hydrogen to W95 NE1, while the O5 ether oxygen atom accepts the hydrogen of W194 NE1, a conserved residue in κ-carrageenases.

Subsites −3 and −2, surprisingly, establish only limited interactions with the κ-carrageenan disaccharide-repeating unit. The sulfate substituent of the D-galactose-4-sulfate residue in subsite −3 is involved in a salt bridge with R151, while the 3,6-anhydro-D-galactose unit in subsite −2 only interacts hydrophobically with Y64. In contrast, the D-galactose-4-sulfate unit at the site of cleavage (−1) is firmly bound to the protein. The O2 and O6 hydroxyl groups form hydrogen bonds with the nucleophile E163 OE1 and W144 NE1, respectively. However, the most crucial interaction with the D-galactose-4-sulfate unit in subsite −1 is probably the strong salt bridge between the sulfate group and arginine R260, by which R260 NH1 and NH2 interact with the O7 and O8 oxygen atoms of the sulfate-ester substituent, respectively. This arginine residue, which is conserved in κ-carrageenases, is itself firmly oriented by a network of hydrogen bonds between other conserved residues (R260 NH1 ↔ L259 O, R260 NH2 and NE ↔ Q171 OE1, Q171 NE2 ↔ D179 OD2, and D179 OD2 ↔ K172 NH). We, therefore, propose that R260 is the main determinant for κ-carrageenan recognition and correctly anchors the polysaccharide chain at the cleavage site.

After the site of cleavage, the protein cavity is particularly narrow. The aglycon binding site is composed of only two subsites. In subsite +1, the 3,6-anhydro-D-galactose unit is surrounded by hydrophobic residues, W266 and F271, located in the inner face of the tunnel-forming β sheet D. Its single free hydroxyl group, at O2, forms a hydrogen bond with the nitrogen atom NE2 of H183, the conserved histidine residue mentioned above as being involved in proton trafficking. In subsite +2, the α face of the D-galactose-4-sulfate hexopyranose ring is stacked against the indole ring of tryptophan W266. At first glance, the sulfate group of this κ-carrageenan monosaccharide unit may be seen as being exposed to the solvant. However, arginine R196, a conserved residue in κ-carrageenases, is close enough to be a potential ligand of this anionic charge. In the native structure of κ-carrageenase reported here, the side chain of R196 is stacked against N269, forming a bidentated hydrogen bond with the O6 hydroxyl group of the D-galactose-4-sulfate unit held in subsite −5. Nevertheless, a 180° torsion of χ3 would be sufficient to place R196 NH1 and NH2 in appropriate positions to ionically interact with the D-galactose-4-sulfate substituent in subsite +2. Given the flexibility of arginine residues, these two possibilities are not mutually exclusive. We suggest that, when the κ-carrageenan chain is bound in a productive manner, R196 establishes a salt bridge with the sulfate group in subsite +2 until the end of the first hydrolysis step. Following aglycon departure, and as subsite +2 becomes empty, the conformation of

R196 then changes, bringing its side chain back to its initial position until the release of the glycosyl-enzyme and the translocation of the κ-carrageenan chain. If this is true, R196 would be a critical residue to account for the processivity of κ-carrageenases.

Potential Roles of the Tunnel in the Processive Endolytic Mechanism of κ-carrageenase

Until now, the only glycoside hydrolases with tunnel-shaped active sites were the cellobiohydrolases of families 6, 7, and 48 [21, 22, 35]. The presence of such tunnels is thought to promote the processive release of cellobiose units from one end of the cellulose chain, while the glucosyl units at the other end remain bound to the protein [21, 33], thus avoiding the diffusion of the polysaccharide chain out of the catalytic cavity between two hydrolytic events. One main inferrence from the structure of *P. carrageenovora* κ-carrageenase is the presence of an active site with a tunnel topology. This finding raises the question of how, in such an endo-processive enzyme, the initial attack occurs. The κ-carrageenase structure is exceptionally stable (average B value = 10.3 Å^2). Yet, the tunnel-forming residues present higher temperature factors. Therefore, as recently demonstrated for *Humicola insolens* [36] and *T. reesei* [37] cellobiohydrolases, the initial endoattack by κ-carrageenase may involve conformational changes opening the tunnel-forming loops.

Following the initial attack, the processively-acting κ-carrageenase must somehow remain associated to its substrate. In processive cellulases such as *T. reesei* cellobiohydrolase I, the cellulose chain is thought to be threaded through the tunnel, via a sliding motion favored by "lubricating" tryptophan residues along the active-site cavity [33]. The dual, heteropolysaccharidic nature of the κ-carrageenan-repeating unit precludes such a sliding mechanism, since the D-galactose-4-sulfate and 3,6-anhydro-D-galactose moieties are accommodated by markedly different subsites (Figure 7). To account for κ-carrageenase processivity, one must thus assume a "hopping" motion involving periodic enlargements of the tunnel, yet with the protein channel serving as a guide along the polysaccharide chain. In retaining hydrolases, hydrolysis can occur only when the glycoside bond at the cleavage site is pointing toward the acid/base catalyst. In carrageenans, the alternation of β-1,4 and α-1,3 linkages results in two successive β-1,4 bonds alternatively pointing up and down, respectively (Figure 1). Assuming that between two rounds of hydrolysis the enzyme-substrate complex cannot undergo rearrangements more extensive than a slight enlarging of the tunnel, the hopping period along the κ-carrageenan would correspond at least to the tetrasaccharide, kappa-neocarratetraose-sulfate. Furthermore, given our model of the κ-carrageenan-κ-carrageenase interactions (Figure 7), the tunnel-forming residues maintain the product side (negative subsites) more than they maintain the aglycon side of the polysaccharide chain. Therefore, the covalent reaction intermediate is likely to be a polysaccharide-enzyme, not a κ-carrageenan oligosaccharide-enzyme. It follows that, in agreement with the threading direction determined for family-7 cellobiohydrolases [21], the protein moves along κ-carrageenan chains from their reducing end to their nonreducing end.

Structural and Evolutionary Relationships within Clan-B of Glycoside Hydrolases

Structural Alignment of Clan-B

The clan-B of glycoside hydrolases encompasses two families only, family-7 and family-16. Family-7 includes cellulases only, namely, endo-1,4-β-glucanases and cellobiohydrolases, which display a fairly good degree of sequence conservation to each other [31]. In contrast, family-16 glycoside hydrolases are quite divergent in their primary structures, and they include polysaccharidases with five different substrate specificities, 1,3–1,4-β-glucanases (lichenases), 1,3-β-glucanases (laminarinases), 1,4-β-endotransglucanases (the plant xyloglucan endotransglycosylases, XETs), as well as 1,3-α-1,4-β-galactanases (κ-carrageenases and agarases) [13].

Three-dimensional structures are now avalailable for three enzyme activities in this clan, i.e., various family-7 cellulases (representative structure PDB code 1cel), two closely related lichenases (representative structure PDB code 2ayh), and one κ-carrageenase (1dyp, this study). The inferred structural sequence alignment of the clan-B of glycoside hydrolases is shown in Figure 8. Only four amino acids are strictly conserved throughout the entire clan. They include the nucleophile and acid/base catalytic residues E163 and E168 (numbering of κ-carrageenase) as well as D165, also thought to be involved with the catalysis. Residues (12) are strictly conserved within family-16. Moreover, 34 amino acids appear to be highly homologous residues. Almost all of these residues are located in the lower part of the β sandwich, between β sheets A and B. Their structural and functional significance is analyzed below.

Folding Landmarks

The conserved residues (60%) belong to the hydrophobic belt framing up the interior of the β sandwich. The main network of contiguous hydrophobic clusters runs along the channel axis, forming a backbone of 25 Å, sustaining the whole structure as follows: F49, V291, V225, W57, L253, A128, F143, Y126, M145, V223, I230, W232, I164, I230, and L182. Two well-conserved minor hydrophobic networks complete the hydrophobic core frame: the first, made up of F49, L81, and L83, is perpendicular to the cleft axis, while the second, close to the external sheet and parallel to the main conserved cluster, is composed of I130, Y221, and V234. Another group of conserved residues with a crucial structural role includes G79, G124, G138, Q171, G236, Q250, and P283 in *P. carrageenovora* κ-carrageenase. All of these residues are located at the beginning of the β strands imposing changes in the direction of the polypeptide chain, either between two antiparallel β strands within the same sheet or between the β stands of two different sheets.

Calcium Binding Site and Other Stabilization Sites

In family-16 proteins, residues G79 and D289 are strictly conserved, and E48 is functionally conserved, whereas these amino acids are absent in family-7. In *B. macerans* 1,3-1,4-β-glucanase, these amino acids are known to constitute a calcium binding site [17]. Chemical unfolding and thermal inactivation experiments have

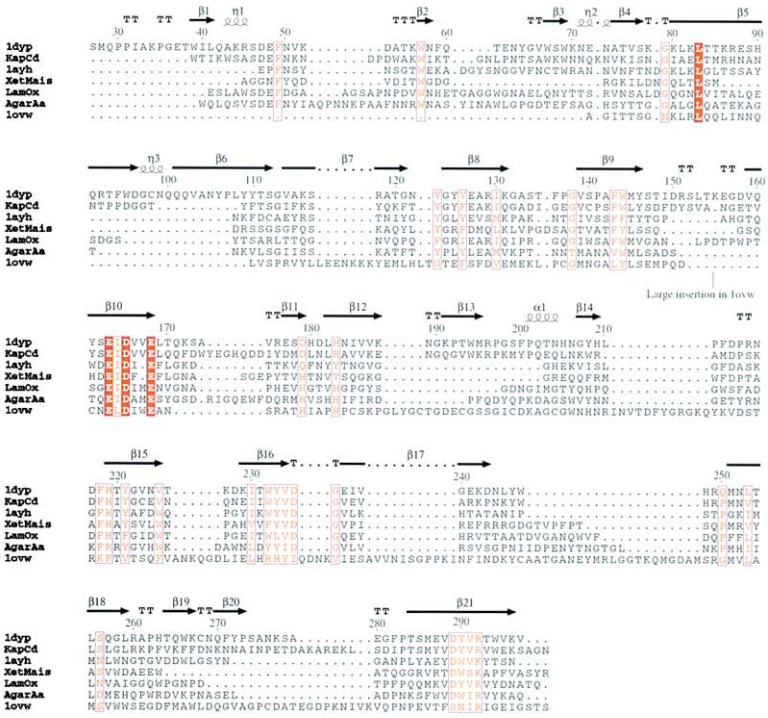

Figure 8. Structure-Based Sequence Alignment of Clan-B of Glycoside Hydrolases

The first line shows the secondary structure elements of κ-carrageenase. The second line displays the κ-carrageenase sequence corresponding to the PDB file 1dyp. The sequences of *B. macerans* 1,3-1,4-β-glucanase (1ayh) and *F. oxysporum* endo-1,4-β-glucanase (1ovw, residues 83–152 and 195–398) have been aligned based on a structural comparison with κ-carrageenase (1dyp) performed with DALI [25]. Sequence alignment of *Zobelia galactanovorans* κ-carrageenase (KapCd, GenBank AF007559, residues 42–310), *Zea mays* xyloglucan-endotransferase (XetMais, GenBank U15781, residues 20–213), *Oerskovia xanthineolytica* 1,3-β-glucanase (LamOx, GenBank U56935, residues 66–306), and *Alteromonas altlantica* β-agarase (AgarAa, GenBank M73783, residues 39–290) was based on the exhaustive alignment of the family-16 glycoside hydrolases previously established by Barbeyron and coworkers [13]. They constitute a representative selection of the different enzymatic groups identified in family-16. The dots represent gaps introduced to improve the alignment. Residues conserved in clan-B are boxed in red. Residues that are conserved or highly homologous in family-16 are shown in red. The figure was prepared with ESPript [58].

shown that binding of calcium at this site stabilizes the 1,3-1,4-β-glucanase structure, prevents aggregation, and possibly enhances thermostability [38]. In the case of κ-carrageenases, the 1dyp structure reveals a cadmium ion in the same location. Since cadmium was used for crystallizing the κ-carrageenase and calcium ions are abundant in sea water (10 mM), it is likely that this κ-carrageenase site genuinely binds calcium, resulting in a higher stability. As these ligands are conserved throughout the family-16, 1,3-β-glucanases, β-agarases, and XETs probably also bind calcium at this location.

On the convex face of the κ-carrageenase, two particular electrostatic interactions are conserved. Near the calcium binding site, OE1 and OE2 of E48 form a bidentate salt bridge with the NH2 and NE, respectively, of R292, while this latter residue is stacked against Y290. As the E48 carbonyl group also binds the cadmium/calcium ion, these interactions probably strengthen the cation binding site. The other conserved salt bridge involves H219 and D235. H219 stacked against F212 makes two hydrogen bonds, one through ND1 and the OD2 carboxylic group of D235, and the other between NE2 and the carbonyl group of N216. It is likely that these two conserved hydrophilic patterns concur with the calcium binding site to stabilize the family-16 proteins.

Clan-B Glycoside Hydrolases Feature an Aromatic Amino Acid at Subsite −1

In spite of the wide substrate diversity of clan-B, all of the glycoside hydrolases of this clan feature one aromatic residue at the site of cleavage, namely, tryptophan in κ-carrageenases, β-agarases, and 1,3-β-glucanases; tyrosine in XETs; and family-7 cellulases and phenylalanine in 1,3-1,4-β-glucanases. In the case of κ-carrageenases, we proposed above that this aromatic residue (W144) is also interacting with the D-galactose-4-sulfate, i.e., the very glycosidic moiety engaged in hydrolysis (Figure 5). Consistently, in the structure of the *F. oxysporum* endoglucanase I complexed with thio-DP5 [26], the lower face of the D-glucose unit in subsite −1 is interacting with Y145, the aromatic residue homologous to W144. Hence, we suggest that, whichever their substrate specificity, clan-B polysaccharidases feature one aromatic residue in subsite −1 of their catalytic machinery, for the correct positioning of the glycoside unit undergoing nucleophilic substitution.

Clan-B Glycoside Hydrolases Fall into Two Subgroups with a Different Active-Site Fine Topology

At strictly conserved positions, all clan-B glycoside hydrolases feature the nucleophile and acid/base catalytic amino acids (E163 and E168, numbering of *P. carrageenovora* κ-carrageenase) as well as D165, a residue that is thought to also be involved in catalysis (see above). However, relative to the fine structure of their active site, clan-B glycoside hydrolases are split into two distinct subgroups (Figure 8). In the first group, which includes 1,3-1,4-β-glucanases and XETs, the catalytic amino acids are held by a regular β strand [17]. The second group, which includes κ-carrageenases (this study), but also β-agarases, 1,3-β-glucanases, as well as family-7 cellulases, is characterized by the occurrence of a β bulge in the homologous β strand, yet with no consequences as to the relative positions of the catalytic amino acids. The κ-carrageenases, β-agarases, 1,3-β-glucanases, and family-7 cellulases again differ from 1,3-1,4-β-glucanases and XETs by the presence of a conserved histidine residue (H183), hydrogen-bound to

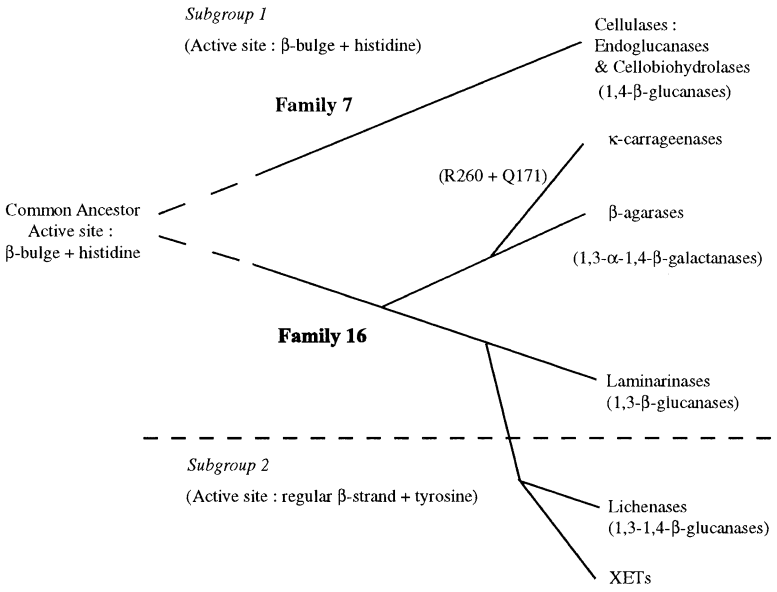

Figure 9. Evolution in Clan-B

Schematic tree based on structural features conserved in each enzyme family.

D165 and thought to be involved with proton trafficking. Instead of this histidine, 1,3-1,4-β-glucanases and XETs feature a tyrosine residue, which does not form hydrogen bonds with any of the catalytic residues [17] and hence plays no role in catalysis.

Evolution of Clan-B Glycoside Hydrolases

The structural differences between agars and κ-carrageenans mainly include the D/L isomery of the α(1,4)-linked galactose residues and the presence or absence of sulfate-ester substituents. Upon phylogenetical analysis of the family-16 of glycoside hydrolases, agarases and κ-carrageenases cluster as two solid, distinct lineages, yet they share a close common ancestor [13]. Phylogenies of red algae inferred from either the plastidial-encoded *rbcL* genes [39] or from the nuclear-encoded 18S rRNA genes [40] indicate that the red algal taxa containing carrageenans in their cell walls have evolved from ancestral agarophytes [41]. It is therefore likely that, as novel polysaccharides appeared, κ-carrageenases emerged from the β-agarases. The differentiation of κ-carrageenases from β-agarases involved the introduction of the basic residue R260, responsible for the binding of D-galactose-4-sulfate. The emergence of κ-carrageenolytic activity also required the substitution at position 171 of a glycine residue that is conserved in all of the other members of family-16 (Figure 8) by glutamine, to ensure the correct positioning of R260.

Our findings on the structural relationships among clan-B of glycoside hydrolases (Figure 8) also indicate that κ-carrageenases, β-agarases, 1,3-β-glucanases, and family-7 cellulases, on the one hand, and 1,3-1,4-β-glucanases and XETs, on the other hand, differ in the fine topology of their active site (β bulge and histidine versus regular β strand and tyrosine). Since the former toplogy is more widely distributed than the latter, including the topology across the family 7-16 boundary, we suggest that it is more ancestral. κ-carrageenases, β-agarases, 1,3-β-glucanases, and family-7 cellulases would, therefore, have conserved a primitive active site, from which evolved the catalytic center of 1,3-1,4-β-glucanases and XETs, following the deletion of one amino acid. This view is consistent with the fact that 1,3-1,4-β-glucanases and XETs are specialized in the hydrolysis of the land-plant polysaccharides lichenans and hemicellulose, respectively, and thus are likely to have emerged from family-16 laminarinases more recently than the other enzymes.

Our current view of the evolution of clan-B glycoside hydrolases, from both the evolutionary relationships of their substrates and the three-dimensional structures available so far, is summarized in Figure 9. From the ancestor of the clan, which featured a β-bulged active center, laminarinases (family-16) and cellulases (family-7) have arisen by gene duplication. While the cellulases (family-7) remained relatively well conserved, family-16 evolved into a large variety of glycoside hydrolases. We propose that the agarases first emerged from the laminarinase branch, which then differentiated to the more recent κ-carrageenases; lichenases and XETs also appear to have emerged relatively recently from the laminarinase branch, upon the differentiation of a novel, non-β-bulged catalytic center.

The Tunnel Active Site of κ-carrageenase: A Feature Involved in the Disruption of κ-carrageenan Double Helices

In addition to their obvious role in enzyme processivity, the tunnels of cellobiohydrolases are involved in lifting cellulose chains from the surface of cellulose crystals and hence prevent their reassociation with crystalline cellulose [33]. The acquisition of a tunnel active site, by loop extension from the open cleft of endoglucanases, thus enhances the degradation of crystalline cellulose, a clear benefit in the evolution of cellulases [21, 22, 42].

In our current view of the endoprocessive mechanism of *P. carrageenovora* κ-carrageenase, the protein initially attacks its substrate as an endoenzyme, with an open tunnel, then translocates toward the nonreducing end of the κ-carrageenan chain, "hopping" by four galactosyl units, probably requiring a slight enlargement of the tunnel cavity. As for cellulases, the natural sub-

strate *P. carrageenovora* κ-carrageenase, which consists of stacked bundles of double-stranded galactan chains networked by more amorphous chains, comprises solid and crystalline regions. It is therefore tempting to speculate that, like in cellobiohydrolases, the tunnel of κ-carrageenase is involved with both processivity and the solid-phase degradation of κ-carrageenan. This latter assumption raises the question of how *P. carrageenovora* κ-carrageenase would disrupt the pseudo-crystalline junction zones of κ-carrageenan fibers and separate the strands of the double helices.

Biological Implications

Carrageenans and agars are gel-forming hydrocolloids extracted from the cell walls of marine red algae. They consist of linear galactans with alternating α-1,3 and β-1,4 linkages and either zero (agarose), one (κ-carrageenan), two (ι-carrageenan), or three (λ-carrageenan) sulfate ester substituents per disaccharide-repeating unit. Their functional properties involve coil-helix transitions leading to the formation of highly ordered aggregates of double-stranded helices [4]. A variety of agar- and carrageenan-degrading enzymes are secreted by marine microbes, which are involved in the diseases of red algae or with recycling their biomass [11].

We have cloned and expressed κ-carrageenases and ι-carrageenases [9, 12, 13, 23, 43]. The κ-carrageenases belong to the family-16 of glycoside hydrolases, a highly divergent group of polysaccharidases, which, together with family-7 cellulases, form clan-B of the glycoside hydrolases [12, 13, 22]. In contrast, ι-carrageenases constitute a novel family that is unrelated to κ-carrageenases [9].

Here, we present the crystal structure of *Pseudoalteromonas carrageenovora* κ-carrageenase, the first structure from a carrageenan-degrading enzyme. It features a tunnel-shaped active site, the first to be found in a noncellulolytic glycoside hydrolase, suggesting that such topologies are associated with the capacity to degrade crystalline polysaccharides, irrespective of their fine chemistry. We identify in the κ-carrageenase cleavage site a unique arginine residue that is involved with κ-carrageenan recognition. We also find that κ-carrageenases exhibit a β bulge-containing catalytic site in contrast to the land-plant polysaccharide-degrading enzymes lichenases and XETs. As red algae emerged as an independent phylum hundreds of million years before the differentiation of terrestrial plants, we argue that agarases and κ-carrageenases arose early in the evolution of clan-B glycoside hydrolases and that the β-bulged catalytic site is the ancestral form in this clan.

Experimental Procedures

Expression and Purification of Native and Selenomethionyl *P. carrageenovora* κ-carrageenase

Full details of the expression and purification of native *P. carrageenovora* κ-carrageenase have been given previously [23]. Briefly, κ-carrageenase was expressed in the pET20b vector (Novagen) as a His-tagged fusion protein in the periplasm of *E. coli* BL21 (DE3) and purified by metal-affinity chromatography on a column of Chelating Fast Flow Sepharose (Pharmacia) loaded with $NiSO_4$. The yield was 4–5 mg/l culture medium, and the purified protein was concentrated to 6 mg/ml using a dialyzing concentrator (Amicon).

The seleno-L-methionine-labeled protein was expressed by a slightly different protocol, since the published procedure using LeMaster medium [44] led to cell lysis and no expression, due to seleno-L-methionine (Se-Met) toxicity. Recombinant *E. coli* B834 (DE3) cells were first incubated at 37°C in a medium similar to that described by Ramakrishnan and coworkers [45] (i.e., M9 medium supplemented with all amino acids [40 mg/l], except methionine; $FeSO_4$ [12.5 mg/l]; and various vitamins) that was supplemented with methionine (40 mg/l), 100 μg/ml ampicillin, and 37 μg/ml chloramphenicol. When the OD_{600} of the culture reached 1.0, the cells were spun down (5000×*g*, 15 min), and the medium was discarded. The cells were resuspended in M9 medium to remove all traces of methionine, then spun down again before being finally resuspended in Ramakrishnan medium supplemented with seleno-L-methionine (10 mg/l), 100 μg/ml ampicillin, and 37 μg/ml chloramphenicol. Expression was then induced for 15 hr at 12°C by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Using this protocol, cell lysis did not occur, and the yield was 1 mg/l. Se-Met-κ-carrageenase was purified in the presence of 250 μM β-mercaptoethanol to prevent oxidation of the seleno-L-methionine without reducing the chelated nickel ion. After purification, the amount of β-mercaptoethanol was increased to 1 mM. The purified Se-Met-κ-carrageenase was concentrated to 6 mg/ml.

Crystallization

Crystallization of Se-Met-κ-carrageenase was first attempted under conditions similar to those already reported [23], except for the addition of 1 mM β-mercaptoethanol. Under these conditions, the protein formed heavy precipitates in half a day, but no crystals were spontaneously produced. However, seeding the drop with small native κ-carrageenase crystals induced the nucleation of Se-Met-κ-carrageenase crystals. These crystals were also used for seeding, to ensure that the final crystals were completely substituted with seleno-L-methionine. The crystals appeared in 24 hr and reached their maximal size in a week. They diffracted as well as native crystals.

Data Collection and Processing

A single crystal was mounted on a loop, transferred to the goniometric head without cryo-additive, and kept at 100K in a nitrogen stream. The MAD data were collected from a Se-Met-κ-carrageenase crystal at three wavelengths around the K absorption edge of selenium (beamline BM 14, ESRF-Grenoble, France). A MAR CCD detector was used. All intensity data were integrated and reduced using DENZO/SCALEPACK software [46]. Data at 1.74 Å resolution at the three wavelengths were input into SOLVE [47] to locate the selenium atoms. The program provided the positions for four of the six atoms expected in the asymmetric unit; the two missing selenium atoms corresponded to M1 and M3 and were assumed to be disordered. The final phases from SOLVE resulted in a figure of merit of 0.68–1.74 Å. Crystallographic statistics for data reduction, phasing, and refinement are reported in Table 1.

Structure Determination and Refinement

The electron density was slightly improved by histogram matching and solvent flattening (35% of solvent) using DM [48], yielding a figure of merit of 0.84–1.74 Å. Due to the high quality of the electron density map, wARP [49] was used to build the model automatically. Thus, 90% and 85% of the main and side chain atoms, respectively, were assigned. The rest of the model was built in the MAD map using O [50]. Refinement was performed using CNS [51], with the λ_{remote} data set to 1.54 Å and using the maximum likelihood target function. The program was set up to automatically compute a cross-validated σ_a estimate [52] and the weighting scheme between the X-ray refinement target and the geometric energy function. Corrections for a flat bulk solvent and for anisotropy in the data were also applied. The σ_a-weighted maps obtained from the subsequent refinement models were used for further model building. The first water molecules were added conservatively to peaks (>2 σ) of the $2F_o$-F_c electron density map that made at least one hydrogen bond with a protein atom or another water molecule. In the final stages,

the sigma cutoff was reduced to 1.5 σ, and water molecules with a B factor $>$ 50 Å were removed. To confirm the proposed Cd binding sites, an anomalous Fourier difference map was computed using amplitude data from a data set collected in the laboratory (λ_{cu} = 1.5418 Å) and phases from the model. The final refined model at 1.54 Å had a crystallographic R_{work} of 17.9% and an R_{free} of 19.3% and contained 271 amino acid residues, 405 water molecules, 7 cadmium ions, and 7 chloride ions.

Modeling Study

In order to investigate the binding of a κ-carrageenan chain to the active site of *P. carrageenovora* κ-carrageenase, κ-neocarrapentaose-sulfate was positioned in the binding channel. The structure of this oligosaccharide is based on the model of the double helix of ι-carrageenan (PDB code 1car) determined by X-ray diffraction on ι-carrageenan fibers [6]. Only one strand was conserved, and the sulfate substituent on each 3-6-anhydro-galactose moiety was replaced by a hydroxyl to generate a single strand of oligo-κ-carrageenan. To correctly position this oligosaccharide, the structure of κ-carrageenase was superimposed on that of *F. oxysporum* endoglucanase I complexed with the nonhydrolysable thiooligosaccharide substrate analog thio-DP5 (PDB code 1ovw) [26] by least square refinement using LSQKAB [53]. In this structure, the pyranose ring at the cleavage point is clearly distorted from the standard $^4C_1$ chair to a $^1S_3$ skew boat conformation with a quasi-axial orientation of the glycosidic bond. Atoms C2, C3, C4, C5, and O5 of a galactose-4-sulfate moiety were superimposed on the equivalent atoms of the distorted glucose at subsite −1. This galactose was then modeled to adopt a skew boat conformation similar to that of thio-DP5. The other residues of κ-neocarrapentaose-sulfate were then manually fitted into the binding tunnel using TURBO [54], optimizing carbohydrate-protein interactions without going too far from the initial stable conformation.

Acknowledgments

We wish to thank Dr. Gordon Leonard for his invaluable help on beamline BM14 (ESRF), Dr. Thierry Vernet for his support on Se-Met-substituted protein expression, and Dr. Richard Kahn and Mr. Eric Girard for the identification of the Cd binding sites.

This work was supported by grants from the Action Concertée Coordonnée Sciences du Vivant (N° V) and GDR 1002 of Centre National de la Recherhe Scientifique (CNRS) "Biology, Biochemistry and Genetics of Marine Algae".

Received : September 28, 2000
Revised : January 5, 2001
Accepted : January 9, 2001

References

1. De Ruiter, G.A., and Rudolph, B. (1997). Carrageenan biotechnology. Trends Food Sci. Technol. *8*, 389–395.
2. Knoll, A.H. (1992). The early evolution of eukaryotes: a geological perspective. Science *256*, 622–627.
3. Kloareg, B., and Quatrano, R.S. (1988). Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. Oceanogr. Mar. Biol. Annu. Rev. *26*, 259–315.
4. Rees, D.A. (1969). Structure, conformation, and mechanism in the formation of polysaccharide gels and networks. Adv. Carbohydr. Chem. Biochem. *24*, 267–332.
5. Anderson, N.S., Campbell, J.W., Harding, M.M., Rees, D.A., and Samuel, J.W. (1969). X-ray diffraction studies of polysaccharide sulphates: double helix models for k- and l-carrageenans. J. Mol. Biol. *45*, 85–99.
6. Arnott, S., Scott, W.E., Rees, D.A., and McNab, C.G. (1974). Iota-carrageenan: molecular structure and packing of polysaccharide double helices in oriented fibers of divalent cation salts. J. Mol. Biol. *90*, 253–267.
7. Morris, E.R., Rees, D.A., and Robinson, G. (1980). Cation-specific aggregation of carrageenan helices: domain model of polymer gel structure. J. Mol. Biol. *138*, 349–362.
8. Potin, P., Sanseau, A., Le Gall, Y., Rochas, C., and Kloareg, B. (1991). Purification and characterization of a new kappa-carrageenase from a marine Cytophaga-like bacterium. Eur. J. Biochem. *201*, 241–247.
9. Barbeyron, T., Michel, G., Potin, P., Henrissat, B., and Kloareg, B. (2000). Iota-carrageenases constitute a novel family of glycoside hydrolases, unrelated to that of kappa-carrageenases. J. Biol. Chem. *275*, 35499–35505.
10. Weigl, J., and Yaphe, W. (1966). The enzymic hydrolysis of carrageenan by *Pseudomonas carrageenovora*: purification of a kappa-carrageenase. Can. J. Microbiol. *12*, 939–947.
11. Bouarab, K., Potin, P., Correa, J., and Kloareg, B. (1999). Sulfated oligosaccharides mediate the interaction between a marine red alga and its green algal pathogenic endophyte. Plant Cell *11*, 1635–1650.
12. Barbeyron, T., Henrissat, B., and Kloareg, B. (1994). The gene encoding the kappa-carrageenase of *Alteromonas carrageenovora* is related to beta-1,3-1,4-glucanases. Gene *139*, 105–109.
13. Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloareg, B. (1998). The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and phylogenetic relationships within family-16 glycoside hydrolases. Mol. Biol. Evol. *15*, 528–537.
14. Malet, C., Jimenez-Barbero, J., Bernabe, M., Brosa, C., and Planas, A. (1993). Stereochemical course and structure of the products of the enzymic action of endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase from *Bacillus licheniformis*. Biochem. J. *296*, 753–758.
15. Potin, P., et al., and Kloareg, B. (1995). Processing and hydrolytic mechanism of the cgkA-encoded kappa-carrageenase of *Alteromonas carrageenovora*. Eur. J. Biochem. *228*, 971–975.
16. Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. *7*, 637–644.
17. Keitel, T., Simon, O., Borriss, R., and Heinemann, U. (1993). Molecular and active-site structure of a Bacillus 1,3-1,4-beta-glucanase. Proc. Natl. Acad. Sci. USA *90*, 5287–5291.
18. Hahn, M., Pons, J., Planas, A., Querol, E., and Heinemann, U. (1995). Crystal structure of *Bacillus licheniformis* 1,3-1,4-beta-D-glucan 4-glucanohydrolase at 1.8 Å resolution. FEBS Lett. *374*, 221–224.
19. Juncosa, M., Pons, J., Dot, T., Querol, E., and Planas, A. (1994). Identification of active site carboxylic residues in *Bacillus licheniformis* 1,3-1,4-beta-D-glucan 4-glucanohydrolase by site-directed mutagenesis. J. Biol. Chem. *269*, 14530–14535.
20. Hahn, M., Olsen, O., Politz, O., Borriss, R., and Heinemann, U. (1995). Crystal structure and site-directed mutagenesis of *Bacillus macerans* endo-1,3-1,4-beta-glucanase. J. Biol. Chem. *270*, 3081–3088.
21. Divne, C., et al., and Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science *265*, 524–528.
22. Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.
23. Michel, G., Barbeyron, T., Flament, D., Vernet, T., Kloareg, B., and Dideberg, O. (1999). Expression, purification, crystallization and preliminary X-ray analysis of the kappa-carrageenase from *Pseudoalteromonas carrageenovora*. Acta Crystallogr. D *55*, 918–920.
24. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.
25. Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.
26. Sulzenbacher, G., Driguez, H., Henrissat, B., Schulein, M., and Davies, G.J. (1996). Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. Biochemistry *35*, 15280–15287.
27. Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. Camb. Phylos. Soc. *28*, 416–436.

28. Sinnott, M.L. (1990). Catalytic mechanisms of glycosyl transfer. Chem. Rev. *90*, 1171–1202.
29. Chan, A.W., Hutchinson, E.G., Harris, D., and Thornton, J.M. (1993). Identification, classification, and analysis of beta-bulges in proteins. Protein Sci. *2*, 1574–1590.
30. Cleland, W.W., Frey, P.A., and Gerlt, J.A. (1998). The low barrier hydrogen bond in enzymatic catalysis. J. Biol. Chem. *273*, 25529–25532.
31. Kleywegt, G.J. *et al.*, and Jones, T.A (1997). The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. *272*, 383–397.
32. Hahn, M., Keitel, T., and Heinemann, U. (1995). Crystal and molecular structure at 0.16 nm resolution of the hybrid Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-M). Eur. J. Biochem. *232*, 849–858.
33. Divne, C., Stahlberg, J., Teeri, T.T., and Jones, T.A. (1998). High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. *275*, 309–325.
34. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K., and Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science *249*, 380–386.
35. Parsiegla, G., et al., and Haser, R. (1998). The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. EMBO J. *17*, 5551–5562.
36. Armand, S., Drouillard, S., Schulein, M., Henrissat, B., and Driguez, H. (1997). A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. J. Biol. Chem. *272*, 2709–2713.
37. Zou, J.Y., et al., and Jones, T.A. (1999). Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Ce16A from *Trichoderma reesei*. Structure *7*, 1035–1045.
38. Keitel, T., Meldgaard, M., and Heinemann, U. (1994). Cation binding to a Bacillus (1,3-1,4)-beta-glucanase. Geometry, affinity and effect on protein stability. Eur. J. Biochem. *222*, 203–214.
39. Freshwater, D.W., Fredericq, S., Butler, B.S., Hommersand, M.H., and Chase, M.W. (1994). A gene phylogeny of the red algae (Rhodophyta) based on plastid rbcL. Proc. Natl. Acad. Sci. USA *91*, 7281–7285.
40. Ragan, M.A., Bird, C.J., Rice, E.L., Gutell, R.R., Murphy, C.A., and Singh, R.K. (1994). A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. Proc. Natl. Acad. Sci. USA *91*, 7276–7280.
41. Fredericq, S., Hommersand, M.H., and Freshwater, D.W. (1996). The molecular systematics of some agar- and carrageenan-containing marine red algae based on *rbc*L sequence analysis. Hydrobiologia *326*, 125–135.
42. Meinke, A., et al., and Gilkes, N.R. (1995). Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. J. Biol. Chem. *270*, 4383–4386.
43. Michel, G., Flament, D., Barbeyron, T., Vernet, T., Kloareg, B., and Dideberg, O. (2000). Expression, purification, crystallization and preliminary X-ray analysis of the iota-carrageenase from *Alteromonas fortis*. Acta Crystallogr. D *56*, 766–768.
44. LeMaster, D.M., and Richards, F.M. (1985). ^{1}H-^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. Biochemistry *24*, 7263–7268.
45. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219–223.
46. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
47. Terwilliger, T.C., and Berendzen, J. (1997). Bayesian MAD phasing. Acta Crystallogr. D *53*, 571–579.
48. Cowtan, K., and Main, P. (1998). Miscellaneous algorithms for density modification. Acta Crystallogr. D *54*, 487–493.
49. Perrakis, A., Sixma, T.K., Wilson, K.S., and Lamzin, V.S. (1997). *wARP*: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Crystallogr. D *53*, 448–455.
50. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. D *47*, 110–119.
51. Brünger, A.T., et al., and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
52. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A *42*, 140–149.
53. CCP4 (1994). The CCP4 (Collaborative Computational Project Number 4) suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.
54. Roussel, A., and Cambillau, C. (1991). TURBO FRODO Program. (Mountain View, CA: Silicon Graphics Geometry Partners Directory).
55. Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. *321*, 557–559.
56. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
57. Nicholls, A. (1992). GRASP: Graphical Representation and Analysis of Surface Properties. (New York*:* Columbia University).
58. Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.
59. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Accession Numbers

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (with the accession codes 1dyp and r1dypsf, respectively) at Rutgers University.